DOCUMENT RESUME

ED 427 708 IR 019 259

AUTHOR Hussam, Ali; Ford, Brian; Ludwig, Gail; Prewitt, Mike; Anderson, Terry

TITLE Semantic Highlighting on the WWW: Educational Implications.

PUB DATE 1998-11-00

NOTE 7p.; In: WebNet 98 World Conference of the WWW, Internet and Intranet Proceedings (3rd, Orlando, FL, November 7-12, 1998); see IR 019 231. Figures may not reproduce clearly.

PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Computer Oriented Programs; Computer Uses in Education; *Electronic Text; Error Patterns; Higher Education; *Information Retrieval; Information Seeking; *Relevance (Information Retrieval); Search Strategies; *Semantics; User Needs (Information); *World Wide Web

IDENTIFIERS *Highlighting (Text); HTML; Search Engines; University of Missouri Columbia

ABSTRACT

The main method of locating information on the World Wide Web is to use a search engine. Given a set of terms, a search engine will return a list of documents containing those terms. Often, though, this list of documents is extremely large. Unfortunately, there are currently no tools to assist the information seeker in determining whether these documents contain desired information or just submitted terms. Two types of search engine errors are possible: false positive errors result from the many connotations which words may convey, and false negative errors result from the many wordings that express similar meanings. To solve these difficulties, a technique called Semantic Highlighting was developed that focuses on meaning rather than terms. This technique enables experts and instructors to highlight the most pertinent portions of documents in a hierarchical manner, allowing students, colleagues, and other users to search more efficiently. It also allows for instructors and experts to assess and communicate directly their assessment of the importance of elements within the documents. Two examples of using Semantic Highlighting as a teaching tool at the University of Missouri-Columbia are included. Five figures present screens of the Semantic Highlighting enhanced search engine output, document visualization, and highlighted documents. (Author/AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 427 708

Semantic Highlighting on the WWW: Educational Implications

Ali Hussam, Brian Ford, Gail Ludwig, Mike Prewitt
615 Locust St. Bldg., University of Missouri, Columbia, MO 65211 USA.
Tel: +1 573 882 5896, E-mail: HussamA@missouri.edu

Terry Anderson
School of Computing and Mathematics, University of Ulster, Newtownabbey, Co. Antrim BT37 0QB
Tel: +44 01232 366903, E-mail: TJ.Anderson@ulst.ac.uk

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract: The main method of locating information on the World Wide Web is to use a search engine. Given a set of terms, a search engine will return a list of documents containing those terms. Often, though, this list of documents is extremely large. Unfortunately, there are currently no tools to assist the information seeker in determining whether these documents contain desired information, or just submitted terms. Two types of search engine errors are possible: false positive errors result from the many connotations which words may convey, and false negative errors result from different wordings that express similar meanings. To solve these difficulties, we focus on meaning rather than terms, developing a technique called Semantic Highlighting. This technique enables experts and instructors to highlight the most pertinent portions of documents in a hierarchical manner, allowing students, colleagues and other users to search more efficiently. It also allows for instructors and experts to assess and communicate directly their assessment of the importance of elements within the documents.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G.H. Marks
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Introduction

Studying the written word is arguably the single most intellectually demanding and time-consuming task of any adult learner. The sustained mental effort of critical thinking, analysis, and interpretation is essential for the learner to develop his/her understanding. Online documents have the great advantage of lower cost and wider availability than paper-based documents. However, in their current forms, online documents appear to be poorer learning tools than their paper-based equivalents. For documents more than a few pages long, users often make a hard copy. This implies that paper documents still offer significant advantages over their onscreen counterparts when it comes to reading significant amounts of non-trivial text. With Semantic Highlighting, we are trying to narrow this gap and enable users to directly manipulate the online documents (in this case the HTML file) without changing the original contents of the document.

The Internet is an important information resource and it will remain so for years to come. Virtually all publicly-accessible data will soon be on it (Metcalfe, 1997). Due to the expansion of the Internet, it will become increasingly more difficult to quickly and effectively locate information.

Many visual information seeking and retrieval methods have been developed to support individuals as they browse, search and mine for data. The search process typically begins with only a broad concept of the details required. Then, as the concept becomes clearer, unwanted data is filtered out and the focus turns to the relevant terms remaining. Finally, the specific details that the search has uncovered are retrieved (Schneiderman, 1997).

On the World Wide Web, this search process usually involves a search engine. Given a set of terms, a search engine will return a list of documents containing those terms. This list is usually ranked according to the total number of hits, or total times all search terms were found, within the document. This system of ranking is often misleading, as it only takes into account the total number of matches without regard for the distribution of those matches among the submitted terms. Also, the list of documents returned is often extremely large. Unfortunately, there are currently no tools available to assist the user in

IR019259

BEST COPY AVAILABLE

determining whether retrieved documents contain desired information, or just submitted terms. Often users must browse many of the returned documents to find relevant data. This is obviously a frustrating and time-consuming process. Two types of search engine errors are possible that cause this phenomena: false positive errors result from the many connotations which words may convey, and false negative errors result from different wordings that express similar meanings. Semantic Highlighting has been developed to focus on meaning rather than terms in a search process.

Why Highlight while Reviewing Documents?

Students often sit down with daunting textbooks and highlighting markers, hoping to flag all significant bits of information. Sometimes they end up with entire pages of highlighted text. Used correctly, the highlighting marker can help emphasize and locate important portions of printed text quickly and easily (Sanders, 1996). Thus, highlighting written text is an important skill to develop. This skill is based on the ability to recognize main ideas and supporting details. In addition, a document could be highlighted in order to outline, classify (customize the information), direct attention, guide, and aid navigation. Until now, highlighting of electronic documents has only been used in a syntactic way, to cause the viewer to notice that a phrase is 'clickable' or 'selected' (Marcus 1992 and Preece et. al. 1994).

When re-reading a document, it is useful to be able to skim the familiar material and focus quickly on the new material. To support this task of efficiently re-reading, a learner typically marks, highlights, and annotates--an important part of processing the information. This activity also generates visual impressions of individual pages, which can be useful in finding particular parts of a document again. Highlighting makes learning quicker and easier, as one can re-read the highlighted parts over and over to learn them. This accords with O'Shea's (1997) call for educational interfaces to "develop effective memory prostheses that support the learner in recalling the fine detail of an increasing volume of electronic interaction". Paper documents always allow the reader to underline, but the support for underlining is often poor or even absent in online documents. Semantic Highlighting offers a solution to this problem.

What is Semantic Highlighting?

The information now available on the Internet pertaining to a particular topic varies greatly in both quantity and quality. The World Wide Web has enabled users to electronically publish information making it easily accessible to millions of people. However, the ability of those people to find relevant material has decreased dramatically as the quantity of information on the Internet grows.

One emerging trend is to enable the user to describe their own material with metadata, "information about data" (Iannella and Waugh, 1997). Warwick (1997) writes that "an element of metadata describes an information resource, or helps provide access to an information resource. Metadata can be use to describe an Internet resource; what it is, what it is about, where it is, and so on." There are three major aspects for the deployment of metadata: description of resources, production of the metadata, and use of the metadata. The key issue is that metadata helps to preserve the contents of the original document. Semantic Highlighting is simply adding a new layer of information (metadata) above the original content layer. The highlighting layer can be removed or modified at any time without interference with the original content layer.

Semantic Highlighting is similar to traditional highlighting, but it is performed on electronic documents, initially those in the Hypertext Markup Language format (HTML) (Hussam et. Al. 1998). Semantic Highlighting allows its users to highlight relevant electronic information directly within a web browser window either manually or automatically through the use of key words. The Semantic Highlighting tools also allow users to view documents highlighted by others, including experts in the field of study addressed by the document. Semantic Highlighting Tools offer users the ability perform the following functions on electronic documents:

- Highlight manually
- Highlight automatically using search strings
- Compare/contrast documents highlighted by different users
- Generate outlines from highlighted content
- Customize highlight colors and categories
- View an entire document's highlighted representation through hierarchical icons

- Save highlighted documents locally or publish them to a server

How Semantic Highlighting Reduces Internet Data Retrieval Time

Semantic Highlighting can reduce the time required to retrieve information on the Internet in two ways. First, Semantic Highlighting can enhance the existing search engine experience, making it quicker and easier for users to find information (Semantic Highlighting Automatic Mode). Secondly, a Semantic Highlighting search domain can be established for documents previously highlighted by experts, allowing users to access pre-classified information (Semantic Highlighting Experts Mode).

In either case, documents retrieved from a search engine can be displayed using the Semantic Highlighting graphical format. This format will allow users to quickly decide which documents contain their desired content. The format will also allow users to rapidly locate that content and immediately see the relations between search terms.

The first hierarchical level of Semantic Highlighting's graphical format adds a pie chart icon and term color-code to standard search engine outputs. By stating the total number of hits each document contains next to a pie chart representing the relative distribution of those hits, users can quickly determine which documents contain the most relevant information. The color-code for each search term is shown in a separate frame below the search engine's output.

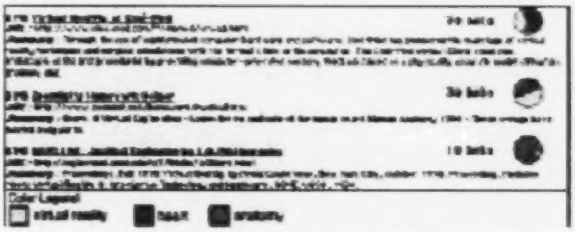

Fig. 1: Semantic Highlighting enhanced search engine output

The second level of Semantic Highlighting can be invoked when a user has determined that a particular document contains the desired information. By 'clicking' on the pie chart icon, the Semantic Highlighting tools will parse the document into standard sixty-line pages. Then, the tools will determine which pages have the highest density of relevant content. These pages will be displayed as thumbnail versions of the full-sized color-coded highlighted pages. The remaining pages will be hierarchically grouped under 'clickable' pie chart icons similar to those in the previous level. This representation will allow users to quickly find the greatest density of relevant content within a document.

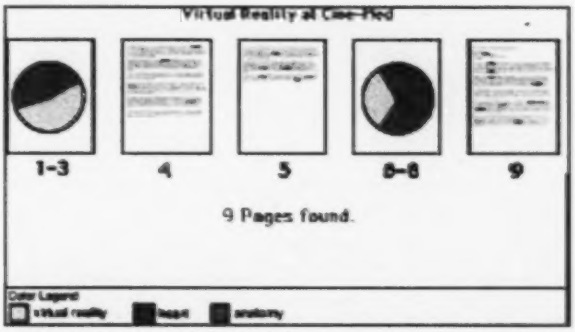

Fig. 2: Semantic Highlighting document visualization

Finally, by 'clicking' on the thumbnail pages, users can retrieve a full-sized version with the color-coded highlights intact.

BEST COPY AVAILABLE

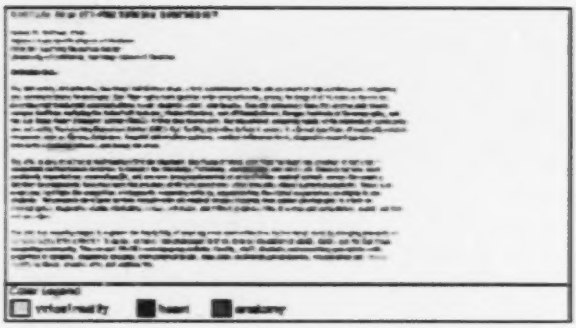

Fig. 3: Semantic Highlighting highlighted document

Semantic Highlighting and Education

In an educational setting, HTML documents that have been analyzed and highlighted by a faculty member or expert may be presented to students. The faculty member or expert will classify the information into pre-defined categories such as main point, major ideas, important terms, etc., based on their knowledge, research, and experience. Students can view their own teachers' highlights and those of other experts. They can also compare and contrast any two sets of highlights. This ability will greatly assist students in understanding the level of importance faculty members and experts place on various pieces of information available on the Web. Using the customization tools, faculty can also generate unique highlighting categories to guide students more efficiently through online class material. This extra guidance will reduce the amount of irrelevant data students retrieve from the Web, provide a condensed set of review material, and help students retain important information (see figures 4 and 5).

Students can also perform their own highlighting on HTML documents. They can then compare their highlights to those of faculty members, experts, or classmates. This comparison will provide students the tools needed to extract important details from a document based on categories defined by the instructor. Also, students can generate outlines from the highlighted material.

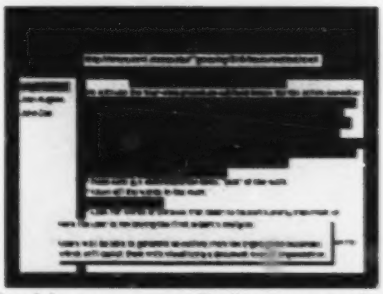

Figure 4: A document is highlighted by an expert. Red represents the main point and green represents examples.

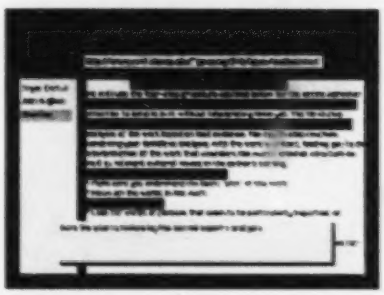

Figure 5: The same document is highlighted by another expert. Red represents the main point and green represents examples.

BEST COPY AVAILABLE

Examples of Using Semantic Highlighting as a Teaching Tool

The following are descriptions of the benefits of Semantic Highlighting in the classroom, provided by faculty members of the University of Missouri-Columbia. Currently these assumptions are based on a conceptual model of Semantic Highlighting. However, by Fall 1998, the faculty will be able to experience these perceived benefits firsthand, as they will be able to use Semantic Highlighting Tools. Several surveys will be conducted to test the validity of Semantic Highlighting use in the classroom, with the results to be published in a future paper. Testing will be conducted simultaneously in Europe and the United States to evaluate Semantic Highlighting in different educational settings.

Dr. Gail S. Ludwig
Department of Geography

A geography student came huffing and puffing to my mapping science class last semester carrying a ten pound notebook filled with pages of web-based documents. I was both amazed and flabbergasted by the thought of all the time, effort and resources (especially paper!) this student expended to capture the information I had linked to my web-based class syllabus. My query as to why the student printed ALL the linked information in the syllabus was met with the standard answer, "Because you might test us on this material".

To give better focus and direction to my students, I downloaded several of the online documents and imported them into a word processing program. Using the bold and italic options, I went through several of the documents identifying the important concepts and ideas contained in the paper. At the end of the class, I felt like the students had a better grasp of identifying key concepts within the paper. The students' next questions were predictable. "Why," the students asked, "couldn't I do this type of highlighting online?" It was a good question and stimulated a great deal of discussion. Why couldn't I highlight the important sections of the web-based documents? Why couldn't I prioritize document content for my students?

The development of Semantic Highlighting as an educational tool to assist teachers, students and general web users to manage the vast amount of information on the Web is a major breakthrough. Like general highlighting done manually in textbooks, it can help identify the key concepts and ideas the instructor feels are important. These highlighted sections will be visible to his/her students logging onto the site. In addition, the students will have the ability to do their own highlighting and compare it with what other students in their class feel is important, or even what other experts in the field identify as important. It will allow a type of collaborative learning to take place on the Web. Although the interaction between students and faculty will not be face-to-face or in real-time, it will allow individuals to work together, examining and evaluating online documents.

Semantic Highlighting is an exciting development that can help educators harness the vast amount of resources on the Web. It can help avoid the information overload that often is experienced when thousands of web sites have information on a specific topic. It is a new tool that educators can add to their technology toolbox to assist them in organizing, prioritizing and understanding the resources available via the World Wide Web.

Dr. Mike Prewitt,
Health Related Sciences
Respiratory Therapy

Keeping up with the medical literature is a strenuous proposition. The stacks of unread journals that collect on desks and in filing cabinets continue to grow larger as our time to read them shrinks. Yet we are pressed to keep current; our patients expect informed practitioners. Students enrolled in health professions programs must, as part of their training, learn to evaluate the strengths and uncover flaws in journal articles. They also must come up with an independent assessment of whether the author's message rings true, and if, in the final analysis, the results are valid. Their task is further complicated by the increasing number of journal articles that contain claims which are tainted by dubious premises, invalid designs, unreliable data, violated assumptions, bias, erroneous methods or

faulty reasoning. The development of Semantic Highlighting, as a tool to assist students, faculty and practitioners in health professions education, would be extremely beneficial in preparing students to evaluate research articles using a practical, critical and efficient approach. Students could compare their journal critiques to those of faculty and other experts in their field, which would provide an effective way to develop skills and ultimately become more efficient at reading health literature.

The Web is changing the way students learn. Computers have become indispensable tools for managing the rapidly growing body of medical information. Semantic Highlighting will become a useful tool in retrieving information in a more efficient and timely manner.

Examples and Future Work

A working example of Semantic Highlighting may be found at: http://pumbaa.atc.missouri.edu/sh.html.

Future work will continue to expand the scope of Semantic Highlighting. Some of the topics currently being considered are:

- Collaborative Semantic Highlighting: This technique will help promote new collaborative learning environments, allowing users to interact in real-time using Semantic Highlighting and chat. Also, these sessions will be more beneficial if a leader or expert is available to facilitate the session.
- Semantic Highlighting Text to Speech: This will allow highlighted text, in outline form, to be read aloud to the user.

References

[Black Academy 1996] Black Academy available at: http://www.jerseyvirtual.net/blacks/bus/mod1a/intro.html, 1996.

[Hussam 1998] Hussam A. et. Al. (1998) Semantic Highlighting . Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, ACM SIGCHI 98, L.A., CA.

[Iannella 1997] Iannella , R., & Waugh, A. (1997) Metadata: Enabling the Internet, Report of CAUSE97, FL, 1997, URL: http://www.cause.org/conference/c97/c97-arc.html

[Marcus 1992] Marcus, A. (1992) Graphic Design for Electronic Documents and User Interfaces, New York: ACM Press.

[Metcalfe 1997] Metcalfe, Bob. (1997) ACMEMBENET, A supplement to communications of the ACM, 40: 5: 8.

[O'Shea 1997] O'Shea T. (1997) A Topology for Educational Interface, Proc. Plenary and Invited Talks,
1997, NY, ACM Press

[Preece 1994] Preece, J. et. al. (1994). Human-Computer Interaction, Wokingham, England: Addison-Wesley.

[Preece 1994] Preece J. (1994)Human-Computer Interaction, Essex: Addison-Wesley

[Sanders 1996] Sanders, Scott. Highlighting and Underlining Written Textbooks available at: http://bvsd.k12.co.us/~ssclass/highlight.html

[Schneiderman 1997] Shneiderman, Ben. (1997). Visual Information Seeking: Advanced Interface and Web Page Design. *HCI International*.

[Warwick 1997] Warwick, C. (1997) Matching Discovery and Recovery, Standards Australia Seminar, national Library of Australia, Australia